ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

PROMIS fatigue, pain intensity, pain interference, pain behavior, physical function, depression, anxiety, and anger scales demonstrate ecological validity

Arthur A. Stone[a,*], Joan E. Broderick[a], Doerte U. Junghaenel[a], Stefan Schneider[a], Joseph E. Schwartz[b]

[a]*USC Dornsife Center for Self-Report Science, University of Southern California, Los Angeles, CA 90089-3332, USA*
[b]*Department of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, NY 11794, USA*

Accepted 21 August 2015; Published online xxxx

Abstract

Objectives: Ecological validity refers to the degree to which instruments faithfully capture information in respondents' natural environments. We examined the ecological validity of eight instruments from the Patient Reported Outcomes Measurement Information System (PROMIS), most of which use 7-day reporting periods, by comparing PROMIS scores with daily diary data as a standard.

Study Design and Setting: Five groups of approximately 100 respondents each completed daily diaries and weekly PROMIS instruments for 4 consecutive weeks: community residents; osteoarthritis patients; women experiencing premenstrual syndrome; men undergoing hernia surgery; and breast cancer patients receiving chemotherapy. The last three groups experienced events (menses, surgery, or chemotherapy, respectively) at standardized times in the protocol to examine symptom changes attributable to these events.

Results: We examined the ability of the PROMIS scales to replicate between-group differences in diaries, to replicate week-to-week changes in diaries, and the correlation between diary and PROMIS scales. As a secondary aim, we examined known-group differences with the PROMIS measures. All three types of ecological validity were strongly confirmed, as was known-group validity for the PROMIS recall scales.

Conclusion: This study adds to the growing literature supporting the reliability and validity of the family of PROMIS instruments.

Keywords: Ecological validity; Patient-reported outcomes; Known-group validity; PROMIS

1. Introduction

Demonstrating the validity of patient-reported outcomes (PROs) is necessary for advancing their use in clinical research and practice. Ecological validity—the degree to which instruments faithfully capture information and experiences in respondents' natural environments—is often overlooked. The purpose of this study was to examine the ecological validity of eight instruments from the Patient Reported Outcomes Measurement Information System (PROMIS) across a range of patient populations.

PROMIS instruments are distinguished from prior PRO measures in many positive ways. The item banks underlying each of the PROMIS instruments were established using principles of modern psychometric theory [item response theory (IRT)] and of qualitative item development. Thousands of items from ''legacy'' instruments were meticulously culled, tested, and edited in a formal process, and new items were written as required [1]. Candidate items were subjected to large-scale testing with thousands of respondents for purposes of calibration and for eliminating problematic items, such as ones with unacceptable differential item functioning. The metric of the PROMIS adult scales is anchored to the mean of the US general population, which enhances the interpretability of the scores. Finally, PROMIS measures are intended to be universally relevant rather than disease specific, that is, the PRO domains are assumed relevant across many clinical populations and levels of severity, allowing for broad

Conflict of interest: A.A.S. declares a potential conflict as Senior Scientist with the Gallup Organization and as a Senior Consultant with ERT, Inc.

* Corresponding author. USC Dornsife Center for Self-Report Science & Department of Psychology, University of Southern California, Los Angeles, CA 90089-3332, USA.

E-mail address: arthur.stone@usc.edu (A.A. Stone).

http://dx.doi.org/10.1016/j.jclinepi.2015.08.029
0895-4356/

What is new?

Key findings

- The findings clearly showed that various associations detected by daily diaries were replicated by the PROMIS recall instruments.

What this adds to what was known?

- PROMIS scales had previously been shown to have many excellent psychometric properties and this study adds ecological validity to those characteristics.

What is the implication and what should change now?

- These results add to the confidence that researchers and clinicians should have using the PROMIS scales.

applicability and comparability across populations and diseases [2].

The item banks generated by the PROMIS researchers allow for customized administration including: very brief short forms; forms designed with sensitivity in certain segments of domain severity; and computerized adaptive tests (CATs), wherein items are dynamically selected in real time according to prior responses [2,3]. In settings where computer administration is possible, CATs are particularly attractive because they provide highly precise scores using few items. Importantly, PROMIS instruments are robust over modes of administration, making them suitable for a wide range of applications [4,5].

The goal of this study is to test the ecological validity of these newly developed PROMIS instruments. Ecological validity is a concept attributed to Egon Brunswik in the 1940s. Its meaning has evolved over the years to indicate the degree to which findings are representative of phenomena that occur in peoples' natural or typical environments, as opposed to atypical or contrived settings [6]. Although usually not considered one of the core psychometric validities, such as discriminant or criterion validity, there is little doubt that researchers and clinicians are increasingly recognizing the importance of using assessments that reflect experience in the "real world." Accurate measurement of everyday behaviors, cognitions, and emotions is a feature that is necessary for developing a sound science, even governmental regulatory agencies have recognized the value of ecological validity. Furthermore, almost all PROMIS instruments ask patients to report over the "past 7 days," one of the most common recall periods in PRO measurement [2]. However, there is concern that retrospective assessments of behavior and internal states may lack ecological validity because people may not be able to accurately or reliably recall experiences over the 7-day reporting period [7–10], potentially reducing the accuracy of these reports.

To address this concern, we compare PROMIS responses using the 7-day recall period to responses collected at the end of each day over the same week. An assumption underlying these analyses is that responses collected with end-of-day diaries more accurately reflect experience by reducing recall bias. Studies comparing data collected with momentary assessment methods (Ecological Momentary Assessment, which prompts respondents in everyday life to report about their immediate experience) to end-of-day diaries support this contention [11], although some slight distortion of end-of-day reports has been found [12]. To compare diary and recall symptoms, we average 7 days of daily ratings to create a weekly score. (We acknowledge that the average is only one of many techniques for creating a weekly summary score [13], but it is straightforward and an intuitive way to do so.) The short recall period (1 day) and the investigator-computed weekly average should provide an accurate measure of weekly symptoms to be compared with 7-day recall.

In this study, we examine the ecological validity of PROMIS CAT measures and short forms by testing the hypothesis that items formatted with 7-day recall provide the same or very similar information to that of a week's set of daily short form ratings that are averaged over the same, 1-week period. If the standard PROMIS instruments using 7-day recall closely track the aggregated daily ratings, then we will conclude that PROMIS recall measures are ecologically valid. The PROMIS set of instruments also includes a single item pain intensity scale, which is very similar to commonly used pain assessments. Although it is not a PROMIS multiitem scale, we included it in this study given the widespread use of the item in medical research.

Ecological validity was examined broadly across a community sample (CS) and four disease samples for eight PROMIS domains. Three distinct hypotheses regarding ecological validity were tested: (1) we hypothesize that differences in mean scale scores among diagnostic groups detected by aggregated daily (short form) ratings will be comparable to group differences in weekly recall measures ("group differences"). This hypothesis is based on the expectation that ecologically valid recall measures will yield diagnostic group differences on domain scales that are similar to those collected on a daily basis. If this was not the case, we would be concerned that cognitive processes of recall were biasing the recall measures in ways that distort group differences. (2) Across all diagnostic groups and within each diagnostic group, we hypothesize moderate to very strong correlations between scores based on daily ratings and those based on weekly recall ("correspondence"); (3) We hypothesize that changes in scale scores from week to week (associated with events occurring during the study period, described below) that